

**CIRCUITS AND METHODS FOR POWER MANAGEMENT IN A
PROCESSOR-BASED SYSTEM AND SYSTEMS USING THE SAME**

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0001] The present invention relates in general to data and signal processing and in particular to circuits and methods for power management in a processor-based system and systems using the same.

DESCRIPTION OF THE RELATED ART

[0002] The power loading of a data processing engine is often a time-dependent function of the software being run and/or the type of data being processed. For example, if the software is partitioned for batch processing, the current draw of the processing engine increases as instructions execution picks-up and decreases as instruction execution drops-off. Similarly, if data are processed in blocks, the current draw increases as each block is processed and decreases during the "quiet" time between blocks. As the load varies, so does the noise injection into the device substrate. Moreover, as the loading changes, signal edge rise and fall times vary as the current available for driving the corresponding capacitive loads varies, (especially for signals generated from the same voltage supply as the processor).

[0003] In almost all applications, the noise caused by a modulated load is undesirable, and in some cases can even be tolerated. In certain applications, however, such as audio decoding and encoding, this noise can significantly impact system performance and adversely affect the quality of the output experienced by the end-user.

SUMMARY OF THE INVENTION

[0004] According to one embodiment of the present inventive concepts, an adaptive clock throttle is disclosed for integrating a clock generator generating a high speed clock and a processing engine operating in response to a processing clock. The adaptive clock throttle generates a plurality of lower speed clocks from the high speed clock, estimates a duty cycle of the processing engine, and selectively gates one of the lower speed clocks to the processing engine as the processing clock to increase the duty cycle of the processing engine.

[0005] According to additional embodiments of these concepts, the adaptive clock throttle is further operable to selectively mask cycles of the selected lower speed clock to adjust the frequency of the processing clock and/or to produce a series of non-periodic clock cycles.

[0006] The total produced effect is to spread the processing of instructions across the entire period of a block without periodicity. This ensures that the electrical load created by the processing engine is relatively constant and without modulation. In turn, substrate noise and/or the modulation of the slope of various signals on the chip is substantially reduced or eliminated.

098299-060203

[0016] FIGURE 7 is a flow chart illustrating the operation of Adaptive Clock Throttle in further detail;

ATTORNEY DOCKET NO.
1081-CA

PATENT

[0017] FIGURE 8 illustrates the typical processing of a block of data in four primary operating regimes.

FD-090-66294860

WSM Docket No.
2836- P128US

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0018] The principles of the present invention and their advantages are best understood by referring to the illustrated embodiment depicted in FIGURE 1 – 8 off the drawings, in which like numbers designate like parts.

[0019] FIGURE 1A is a general overview of an audio information decoder 100 embodying the principles of the present invention. Decoder 100 is operable to receive data in any one of a number of formats, including compressed data conforming to the AC-3 digital audio compression standard, (as defined by the United States Advanced Television System Committee) through a compressed data input port CDI. An independent digital audio data (DAI) port provides for the input of PCM, S/PDIF, or non-compressed digital audio data.

[0020] A digital audio output (DAO) port provides for the output of multiple-channel decompressed digital audio data. Independently, decoder 100 can transmit data in the S/PDIF (Sony-Phillips Digital Interface) format through transmit port XMT.

[0021] Decoder 100 operates under the control of a host microprocessor through a host port HOST and supports debugging by an external debugging system through the debug port DEBUG. The CLK port supports the input of a master clock for generation of the timing signals within decoder 100.

[0022] While decoder 100 can be used to decompress other types of compressed digital data, it is particularly advantageous to use decoder 100 for decompression of AC-3 Bitstreams. Therefore, for understanding the utility and

[0024] Next, the exponents are encoded into coarse representation of the signal spectrum (spectral envelope), which is in turn used in a bit allocation routine that determines the number of bits required to encoding each mantissa. The spectral envelope and the coarsely quantized mantissas for six audio blocks (1536 audio samples) are formatted into an AC-3 frame. An AC bit stream is a sequence of the AC-3 frames.

WSM Docket No.
2836- P128US

09796-08-01

[0027] FIGURE 1B shows decoder 100 embodied in a representative system 103. Decoder 100 as shown includes three compressed data input (CDI)

pins for receiving compressed data from a compressed audio data source 104 and an additional three digital audio input (DAI) pins for receiving serial digital audio data from a digital audio source 105. Examples of compressed serial digital audio source 105, and in particular of AC-3 compressed digital sources, are digital video discs and laser disc players.

[0028] Host port (HOST) allows coupling to a host processor 106, which is generally a microcontroller or microprocessor that maintains control over the audio system 103. For instance, in one embodiment, host processor 106 is the microprocessor in a personal computer (PC) and System 103 is a PC-based sound system. In another embodiment, host processor 106 is a microcontroller in an audio receiver or controller unit and system 103 is a non-PC-based entertainment system such as conventional home entertainment systems produced by Sony, Pioneer, and others. A master clock, shown here, is generated externally by clock source 107. The debug port (DEBUG) consists of two lines for connection with an external debugger, which is typically a PC-based device.

[0029] Decoder 100 has six output lines for outputting multi-channel audio digital data (DAO) to digital audio receiver 109 in any one of a number of formats including 3-lines out, 2/2/2, 4/2/0, 4/0/2 and 6/0/0. A transmit port (XMT) allows for the transmission of S/PDIF data to an S/PDIF receiver 110. These outputs may be coupled, for example, to digital to analog converters or codecs for transmission to analog receiver circuitry.

[0030] FIGURE 1C is a high level functional block diagram of a multichannel audio decoder 100 embodying the principles of the present invention. Decoder 100 is divided into two major sections, a Processor Block 101 and the I/O Block 102. Processor Block 106 includes two digital signal processor (DSP) cores, DSP memory, and system reset control. I/O Block 102 includes inter-processor communication registers, peripheral I/O units with their necessary support logic, and interrupt controls. Blocks 101 and 102 communicate via interconnection with the I/O buses of the respective DSP cores. For instance, I/O Block 102 can generate interrupt requests and flag information for communication with Processor Block 101. All peripheral control and status registers are mapped to the DSP I/O buses for configuration by the DSPs.

[0031] FIGURE 2 is a detailed functional block diagram of processor block 101. Processor block 101 includes two DSP cores 200a and 200b, labeled DSPA and DSPB respectively. Cores 200a and 200b operate in conjunction with respective dedicated program RAM 201a and 201b, program ROM 202a and 202b, and data RAM 203a and 203b. Shared data RAM 204, which the DSPs 200a and 200b can both access, provides for the exchange of data, such as PCM data and processing coefficients, between processors 200a and 200b. Processor block 101 also contains a RAM repair unit 205 that can repair a predetermined number of RAM locations within the on-chip RAM arrays to increase die yield.

[0032] DSP cores 200a and 200b respectively communicate with the peripherals through I/O Block 102 via their respective I/O buses 206a, 206b. The

peripherals send interrupt and flag information back to the processor block via interrupt interfaces 207a, 207b.

[0033] FIGURE 3 is a detailed functional block diagram of I/O block 102. Generally, I/O block 102 contains peripherals for data input, data output, communications, and control. Input Data Unit 1300 accepts either compressed analog data or digital audio in any one of several input formats (from either the CDI or DAI ports). Serial/parallel host interface 1301 allows an external controller to communicate with decoder 100 through the HOST port. Data received at the host interface port 1301 can also be routed to input data unit 1300.

[0034] IPC (Inter-processor Communication) registers 1302 support a control-messaging protocol for communication between processing cores 200 over a relatively low-bandwidth communication channel. High-bandwidth data can be passed between cores 200 via shared memory 204 in processor block 101.

[0035] Clock manager 1303 is a programmable PLL/clock synthesizer that generates common audio clock rates from any selected one of a number of common input clock rates through the CLKIN port. Clock manager 1303 includes an STC counter which generates time information used by processor block 101 for managing playback and synchronization tasks. Clock manager 1303 also includes a programmable timer to generate periodic interrupts to processor block 101.

[0038] In general, I/O registers are visible on both I/O buses, allowing access by either DSPA (200a) or DSPB (200b). Any read or write conflicts are resolved by treating DSPB as the master and ignoring DSPA.

[0040] In a dual-processor environment like decoder 100, it is important to partition the software application optimally between the two processors 200a, 200b to maximize processor usage and minimize inter-processor communication. For this, the dependencies and scheduling of the tasks of each processor must be analyzed. The algorithm must be partitioned such that one processor does not unduly wait for the other and later be forced to catch up with pending tasks. For example, in most audio decompression tasks including Dolby AC-3®, the

[illegible]

[0042] For example, Dolby AC-3® requires 20-bit resolution PCM output which corresponds to 120dB of dynamic range. The decoder uses a 24-bit DSP which incorporates rounding, saturation and 48-bit accumulators in order to achieve the desired 20-bit precision. In addition, analog performance should at least preserve 95dB S/N and have a frequency response of +/- 0.5dB from 3 Hz to 20 kHz.

WSM Docket No.
2836- P128US

implemented such that it can be programmed by software to implement different applications. The firmware is the fixed portion of software portion including the boot loader, other fixed function code and ROM tables. Since such a system can be programmed, it is advantageously flexible and has less hardware risk due to simpler hardware demands.

[0044] There are several benefits to the dual core (DSP) approach according to the principles of the present invention. DSP cores 200A and 200B can work in parallel, executing different portions of an algorithm and increasing the available processing bandwidth by almost 100%. Efficiency improvement depends on the application itself. The important thing in the software management is correct scheduling, so that the DSP engines 200A and 200B are not waiting for each other. The best utilization of all system resources can be achieved if the application is of such a nature that can be distributed to execute in parallel on two engines. Fortunately, most of the audio compression algorithms fall into this category, since they involve a transform coding followed by fairly complex bit allocation routine at the encoder. On the decoder side the inverse is done. Firstly, the bit allocation is recovered and the inverse transform is performed. This naturally leads into a very nice split of the decompression algorithm. The first DSP core (DSPA) works on parsing the input bitstream, recovering all data fields, computing bit allocation and passing the frequency domain transform coefficients to the second DSP (DSPB), which completes the task by performing the inverse transform (IFFT or IDCT depending on the algorithm). While the second DSP is finishing the transform for a channel n, the

first DSP is working on the channel $n+1$, making the processing parallel and pipelined. The tasks are overlapping in time and as long as tasks are of similar complexity, there will be no waiting on either DSP side. Once the transform for each channel is completed, DSPB can postprocess this PCM data according to the desired algorithm, which could include digital tone control.

[0045] Decoder 100, as discussed above, includes shared memory of 544 words as well as communication "mailbox" (IPC block 1302) consisting of 10 I/O registers (5 for each direction of communication). FIGURE 4 is a diagram representing the shared memory space and IPC registers (1302).

[0046] One set of communication registers looks like this:

- (a) AB_command_register (DSPA write/read, DSPB read only)
- (b) AB_parameter1_register (DSPA write/read, DSPB read only)
- (c) AB_parameter2_register (DSPA write/read, DSPB read only)
- (d) AB_message_semaphores (DSPA write/read, DSPB write/read as well)
- (e) AB_shared_memory_semaphores (DSPA write/read, DSP B read only) where AB denotes the registers for communication from DSPA to DSPB.

[0047] Similarly, the BA set of registers are used in the same manner, with simply DSPB being primarily the controlling processor.

[0048] Shared memory 204 is used as a high throughput channel, while communication registers serve as low bandwidth channel, as well as semaphore variables for protecting the shared resources.

[0049] Both DSPA and DSPA 200a, 200b can write to or read from shared memory 204. However, software management provides that the two DSPs never write to or read from shared memory in the same clock cycle. It is possible, however, that one DSP writes and the other reads from shared memory at the same time, given a two-phase clock in the DSP core. This way several virtual channels of communications could be created through shared memory. For example, one virtual channel is transfer of frequency domain coefficients of AC-3 stream and another virtual channel is transfer of PCM data independently of AC-3. While DSPA is putting the PCM data into shared memory, DSPB might be reading the AC-3 data at the same time. In this case both virtual channels have their own semaphore variables which reside in the AB_shared_memory_semaphores registers and also different physical portions of shared memory are dedicated to the two data channels. AB_command_register is connected to the interrupt logic so that any write access to that register by DSPA results in an interrupt being generated on the DSP B, if enabled. In general, I/O registers are designed to be written by one DSP and read by another. The only exception is AB_message_semaphore register which can be written by both DSPs. Full symmetry in communication is provided even though for most applications the data flow is from DSPA to DSP B. However, messages usually flow in either direction, another set of 5 registers are provided as shown in FIGURE 4 with BA prefix, for communication from DSPB to DSPA.

09876 E-4 : 0E0703L

WSM Docket No.
2836- P128US

it must check that the previous set of coefficients, for the previous channel has been taken by the DSPB. While DSPA is polling the semaphore bit which is in AB_shared_memory_semaphore register it may receive a message from DSPB, via interrupt, that the coefficients are taken. In this case DSPA resets the semaphore bit in the register in its interrupt handler. This way DSPA has an exclusive write access to the AB_shared_memory_semaphore register, while DSPB can only read from it. In case of AC-3, DSPB is polling for the availability of data in shared memory in its main loop, because the dynamics of the decode process is data driven. In other words there is no need to interrupt DSPB with the message that the data is ready, since at that point DSPB may not be able to take it anyway, since it is busy finishing the previous channel. Once DSPB is ready to take the next channel it will ask for it. Basically, data cannot be pushed to DSPB, it must be pulled from the shared memory by DSPB.

[0052] The exclusive write access to the AB_shared_memory_semaphore register by DSPA is all that more important if there is another virtual channel (PCM data) implemented. In this case, DSPA might be putting the PCM data into shared memory while DSPB is taking AC-3 data from it. So, if DSPB was to set the flag to zero, for the AC-3 channel, and DSPA was to set PCM flag to 1 there would be an access collision and system failure will result. For this reason, DSPB is simply sending message that it took the data from shared memory and DSPA is setting shared memory flags to zero in its interrupt handler. This way full synchronization is achieved and no access violations performed.

[0053] For a complete description of exemplary decoder 100 and its advantages, reference is now made to co-assigned U.S. Patent No. 6,081,783 entitled "DIGITAL AUDIO DECODING CIRCUITRY, METHODS AND SYSTEMS".

[0054] In the case of audio encoding / decoding, data are transmitted and processed as blocks defined by standard protocols. For example, for PCM data, these blocks consist of 16 audio samples, for AC3, 256 samples, and for MPEG Layer 3, 576 samples. As each block is processed, the current draw of the processing engine correspondingly increases as generally illustrated by the solid line in FIGURE 5. The dashed-line in this figure corresponds to the power supply droop, which is roughly the inverse of the power supply current draw.

[0055] As is apparent from FIGURE 5, the noise injected into the substrate as the current load modulates will also be periodic. This noise, as well as its harmonics, can fall within the audible band of the audio output. This is especially true in audio systems in which the digital to analog converters (DACs) driving the audio output are integrated onto the same chip as the processing engine and therefore subject to the same substrate noise injection.

[0056] More significantly, as the loading changes with each block of data, the slope of the edges of the clocks used by the DACs, or in the case of an off-chip DAC, used to transmit data to the DACs, modulates. The resulting change in the clock switching points directly contributes to the output noise floor. This is particularly true with respects to changes in the MCLK switching point which drives the DACs. (For a standard serial interface to an off-chip device, the

master clock (MCLK), along with the left-right channel clock (LRCLK) and the serial clock (SCLK) are used to time the data transfer and are subject to the same triggering-edge slope modulation).

[0057] One possible approach to solving the problems discussed above would be to force the processor to operate continuously, even if the tasks being performed were not strictly necessary to the overall processing operations. For example, the quiet times between processing blocks could be used for the performance of non-critical or even dummy operations such that the power consumption would remain relatively constant. The noise floor would generally increase, but there would be substantially fewer spikes and periodic noise. The more continuous noise is less obviously audible and may be tolerable. However, the primary disadvantage to this approach is power consumption; since the processing engine is constantly executing tasks, it is also constantly consuming power.

[0058] Another approach would be to lower the processor clock frequency until the processing of the data is relatively continuous with respect to the data input rate. For example, consider an audio application where the sampling frequency (f_s) is 48 kHz and the nominal processing rate of the processor is 49.152 MIPS. Under these conditions, the processor can execute up to 1024 instructions per sample. Notwithstanding, if the given application only requires, for example, 740 instructions per sample, then the processor clock could be reset to a new clock frequency of:

$$740 / 1024 * 49.152 \text{ MIPS} \approx 35.5 \text{ MIPS}$$

This approach however is difficult to implement, since typically the PLL generating the processor clock, also generates the clocks used in the DAC and/or to transmit data to an external DAC for conversion. It also does not address the problem of a rapidly changing load.

[0059] FIGURE 6 generally depicts a system including an Adaptive Clock Throttle 600 embodying the inventive concepts interfacing conventional clock generation circuitry 601 and an associated data processing engine 602 which could be, for example, a digital signal processor (DSP), microprocessor, microcontroller or other circuit block processing data in response to a clock. In the specific example of system 100, clock throttle 600 can be used in conjunction with either DSPA or DSPB and disposed in clock manager block 1303. The operation of Adaptive Clock Throttle is described in further detail in the flow chart of FIGURE 7.

[0060] According to the principles of the present invention, the processor clock frequency is modulated on an instruction by instruction block basis. In other words, the processor clock rate is varied according to need, as discussed in detail below. In the exemplary case of system 100, the processor clock driving either DSPA, DSPB or both can be varied; the inventive principles are not limited to two DSP devices, or even DSPs.

[0061] For purposes of the following discussion, a DSP clock frequency of 49 MHz will be assumed for convenience, although the principles of the present

invention are also not limited to this clock frequency. In addition to the DSP clock, a high speed clock with a substantially higher frequency than the DPS clock will be generated on-chip. In the present example, the high speed clock will be assumed to have a frequency of 98 MHz, or 2 times the DSP clock frequency of 49 MHz. This is represented in FIGURE 7 at Step 701.

[0062] Generally, the high speed clock is selected to provide a number of clock edges suitable to generate clock "building blocks" of a desired resolution. These clock building blocks can be generated by dividing down the high speed clock by selected divide ratios (for symmetric clock building blocks) or by selectively gating the high speed clock (for symmetric or asymmetric clock building blocks) at Step 702. For example, from a 98 MHz high speed clock, the following symmetric clocks can be generated by frequency division:

TABLE 1

Divisor	Frequency	Period
/1	98.0 MHz	10.18 nsec
/2	49.0 MHz	20.35 nsec
/3	32.67 MHz	30.52 nsec
/4	24.50 MHz	40.72 nsec

[0063] These clocks can then be used to selectively clock the processing engine. First consider the steady state case where PCM data are streamed in

16-sample blocks at a sampling frequency of 48 kHz with the processing engine processing those data at a constant rate of 740 instructions per sample. In this case, the number of clock periods (active edges), assuming 1 clock cycle to execute 1 instruction, of the high speed (98 MHz) per block is:

$$(1) \quad (740 \text{ clock cycles / sample}) * (16 \text{ samples / block}) = \\ 11840 \text{ clock cycles/ block}$$

Moreover, the active period of 1 PCM block :

$$(2) \quad (16 \text{ samples}) * (48 \text{ k samples/ sec})^{-1} = 333 \text{ } \mu\text{sec}$$

Thus, the processing engine clock must be sufficient to provide 11840 clock cycles in 333 μ sec.

[0064] Given the set of clock building blocks of Table 1, there are number of different combinations that will generate 11840 clock cycles in 333 μ sec. One example, in which 98.0 MHz and 24.50 MHz clocks are used, is represented by the equation:

$$(3) \quad (M \text{ samples} * 10.18 \text{ nsec}) + (N \text{ samples} * 40.74 \text{ nsec}) = \\ 333 \text{ } \mu\text{sec,}$$

where:

$$(4) \quad M + N = 11840 \text{ samples for PCM.}$$

[0065] In actual practice, the instruction loading is not constant, as the data type changes and/or the processing requirements change. Additionally, the clock building blocks are preferably selected and gated to avoid periodic modulation of the electrical load and the consequent creation of noise in the signal output. The inventive principles advantageously address both of these problems.

[0066] During the processing of a block of data, a processing engine typically will execute the necessary instructions and then transition into a sleep mode until the next data are ready and available for processing. FIGURE 8 illustrates the typical processing of a block of data in four primary operating regimes. During time Period 1, the processor is in a sleep mode and waiting for data. The processor comes out of sleep during Period 2, receives the new data and executes the instructions necessary for a given task. After the task is complete, the processor returns to sleep during Period 3 while waiting for output space to become available. The processor is active again during Period 4 for outputting the data. In this case, the total duty cycle is:

$$(5) \quad \text{Duty cycle} = (N_2 + N_4) / (N_1 + N_2 + N_3 + N_4)$$

where N_1 , N_2 , N_3 and N_4 are the number of processor clock cycles in operating regimes 1-4 respectively. This process then repeats for the next block of data.

[0070] The Target Speed (frequency) for the processor clock is then calculated at Step 705 as :

$$(6) \quad \text{Target Speed} = (\text{Target Duty Cycle}) * (\text{Current speed})$$

[0071] The processor clock frequency is updated for the current 256 clock cycles at Steps 706-709. Continuing with the assumption that the target duty cycle is 37%, (i.e. target speed < current speed) this is preferably implemented as follows. Aside from overhead (which can easily be added to the calculation by increasing the duty cycle percentage for example), 37% of the maximum processor clock frequency of 98 MHz is approximately 36.6 MHz and thus a clock speed throttle-back is necessary. The new frequency can be generated by taking the single clock building block of the next highest frequency, in this case the 49.0 MHz frequency from the table, and then selectively masking-off (gating) clock cycles until the frequency of the active edges is approximately 36.6 MHz.

[0072] The process of masking off clock cycles can follow a selected pattern (preferably non-periodic) or can be done at random in response to a pseudorandomly generated gating signal. For example, a gaussian (white noise) distribution can be sampled to select the cycles to be masked. In either case, the goal is to eliminate periodicity in the clock and thus periodicity in the modulation of the load. The total produced effect is to spread the processing of instructions across the entire 333 μ sec period for the block without periodicity, such that the electrical load created by the processor is relatively constant and without modulation.

[0073] Dynamic changes in the processing engine load, for example due to a change in the data being processed or the number of instructions per

instruction block, are compensated for by changing the clock duty cycle and/or changing the clock building blocks through a feedback loop in adaptive clock throttle 600.

[0074] The first case to consider is when the loading decreases and the processor clock must be throttled-back at decision block 706. Here:

- (8) $\text{New speed} = (\text{Target Speed} - \text{Current Speed}) * \alpha +$

Current Speed

where α is a factor selected to insure that the feedback loop adjusts the clock frequency in small steps for stability. The new clock is then generated by either masking additional active edges from the current clock building block, or if the new clock frequency is below a lower clock building block, by selecting that lower frequency clock building block and then masking edges as described above.

[0075] The second case is when the loading increases (target speed > current speed at Step 706) and the processor clock frequency must be increased to meet the new demand. In the preferred embodiment, a dipstick is used to determine at Step 710 whether the processor data buffers are in a state of "soft" underflow or a state of "hard" underflow as a result in the increase in the data input rate. The dipstick calculations for the data buffer are:

- (9) Dipstick = (Write pointer – Read pointer) mod size = Fullness;

and

- (10) $\text{Emptiness} = (\text{Size of Buffer} - \text{Dipstick})$

where reads from the buffer are typically done on a sample by sample basis,

[0078] Although the present invention and its advantages have been described in detail, it should be understood that various changes, substitutions and alterations can be made herein without departing from the spirit and scope of the invention as defined by the appended claims.